

# FAST MONOTONE SUMMATION OVER DISJOINT SETS\*

PETTERI KASKI<sup>1</sup>, MIKKO KOIVISTO<sup>2</sup>, AND JANNE H. KORHONEN<sup>2</sup>

**ABSTRACT.** We study the problem of computing an ensemble of multiple sums where the summands in each sum are indexed by subsets of size  $p$  of an  $n$ -element ground set. More precisely, the task is to compute, for each subset of size  $q$  of the ground set, the sum over the values of all subsets of size  $p$  that are *disjoint* from the subset of size  $q$ . We present an arithmetic circuit that, without subtraction, solves the problem using  $O((n^p + n^q) \log n)$  arithmetic gates, all monotone; for constant  $p, q$  this is within the factor  $\log n$  of the optimal. The circuit design is based on viewing the summation as a “set nucleation” task and using a tree-projection approach to implement the nucleation. Applications include improved algorithms for counting heaviest  $k$ -paths in a weighted graph, computing permanents of rectangular matrices, and dynamic feature selection in machine learning.

## 1. INTRODUCTION

**1.1. Weak algebraisation.** Many hard combinatorial problems benefit from *algebraisation*, where the problem to be solved is cast in algebraic terms as the task of evaluating a particular expression or function over a suitably rich algebraic structure, such as a multivariate polynomial ring over a finite field. Recent advances in this direction include improved algorithms for the  $k$ -path [26], Hamiltonian path [4],  $k$ -coloring [9], Tutte polynomial [6], knapsack [22], and connectivity [14] problems. A key ingredient in all of these advances is the exploitation of an algebraic catalyst, such as the existence of additive inverses for inclusion–exclusion, or the existence of roots of unity for evaluation/interpolation, to obtain fast evaluation algorithms.

Such advances withstanding, it is a basic question whether the catalyst is *necessary* to obtain speedup. For example, fast algorithms for matrix multiplication [11, 13] (and combinatorially related tasks such as finding a triangle in a graph [1, 17]) rely on the assumption that the scalars have a ring structure, which prompts the question whether a weaker structure, such as a semiring without additive inverses, would still enable fast multiplication. The answer to this particular question is known to be negative [19], but for

---

<sup>1</sup>HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY HIIT AND DEPARTMENT OF INFORMATION AND COMPUTER SCIENCE, AALTO UNIVERSITY, FINLAND

<sup>2</sup>HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY HIIT AND DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF HELSINKI, FINLAND

*E-mail addresses:* `petteri.kaski@aalto.fi`, `mikko.koivisto@cs.helsinki.fi`, `janne.h.korhonen@cs.helsinki.fi`.

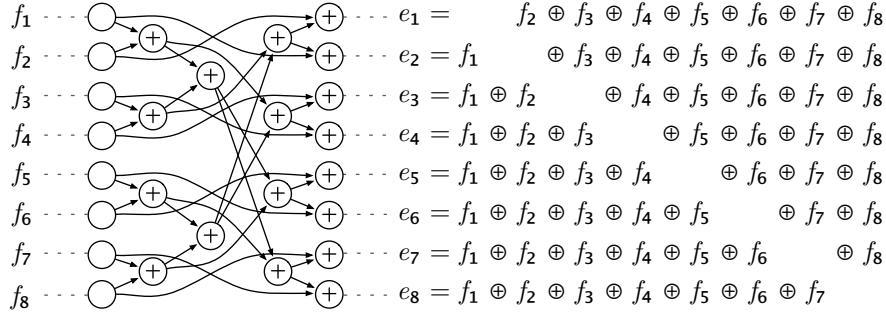
\*This paper is an extended version of a conference abstract by the present authors [18]. The research was supported in part by the Academy of Finland, Grants 252083 (P.K.), 256287 (P.K.), and 125637 (M.K.), and by the Helsinki Doctoral Programme in Computer Science - Advanced Computing and Intelligent Systems (J.K.).

many of the recent advances such an analysis has not been carried out. In particular, many of the recent algebraisations have significant combinatorial structure, which gives hope for *positive* results even if algebraic catalysts are lacking. The objective of this paper is to present one such positive result by deploying *combinatorial* tools.

**1.2. A lemma of Valiant.** Our present study stems from a technical lemma of Valiant [23] encountered in the study of circuit complexity over a monotone versus a universal basis. More specifically, starting from  $n$  variables  $f_1, f_2, \dots, f_n$ , the objective is to use as few arithmetic operations as possible to compute the  $n$  sums of variables where the  $j$ th sum  $e_j$  includes all the other variables except the variable  $f_j$ , where  $j = 1, 2, \dots, n$ .

If additive inverses are available, a solution using  $O(n)$  arithmetic operations is immediate: first take the sum of all the  $n$  variables, and then for  $j = 1, 2, \dots, n$  compute  $e_j$  by subtracting the variable  $f_j$ .

Valiant [23] showed that  $O(n)$  operations suffice also when additive inverses are *not* available; we display Valiant's elegant combinatorial solution for  $n = 8$  below as an arithmetic circuit. [[Please see Appendix A for the general case.]]



**1.3. Generalising to higher dimensions.** This paper generalises Valiant's lemma to higher dimensions using purely combinatorial tools. Accordingly, we assume that only very limited algebraic structure is available in the form of a commutative semigroup  $(S, \oplus)$ . That is,  $\oplus$  satisfies the associative law  $x \oplus (y \oplus z) = (x \oplus y) \oplus z$  and the commutative law  $x \oplus y = y \oplus x$  for all  $x, y, z \in S$ , but nothing else is assumed.

By “higher dimensions” we refer to the input not consisting of  $n$  values (“variables” in the example above) in  $S$ , but rather  $\binom{n}{p}$  values  $f(X) \in S$  indexed by the  $p$ -subsets  $X$  of  $[n] = \{1, 2, \dots, n\}$ . Accordingly, we also allow the output to have higher dimension. That is, given as input a function  $f$  from the  $p$ -subsets  $[n]$  to the set  $S$ , the task is to output the function  $e$  defined for each  $q$ -subset  $Y$  of  $[n]$  by

$$(1) \quad e(Y) = \bigoplus_{X: X \cap Y = \emptyset} f(X),$$

where the sum is over all  $p$ -subsets  $X$  of  $[n]$  satisfying the intersection constraint. Let us call this problem  $(p, q)$ -disjoint summation.

In analogy with Valiant's solution for the case  $p = q = 1$  depicted above, an algorithm that solves the  $(p, q)$ -disjoint summation problem can now be viewed as a circuit consisting of two types of gates: *input gates* indexed by  $p$ -subsets  $X$  and *arithmetic gates* that perform the operation  $\oplus$ , with certain

arithmetic gates designated as output gates indexed by  $q$ -subsets  $Y$ . We would like a circuit that has as few gates as possible. In particular, does there exist a circuit whose size for constant  $p, q$  is within a logarithmic factor of the lower bound  $\Theta(n^p + n^q)$ ?

**1.4. Main result.** In this paper we answer the question in the affirmative. Specifically, we show that a circuit of size  $O((n^p + n^q) \log n)$  exists to compute  $e$  from  $f$  over an arbitrary commutative semigroup  $(S, \oplus)$ , and moreover, there is an algorithm that constructs the circuit in time  $O((p^2 + q^2)(n^p + n^q) \log^3 n)$ . These bounds hold uniformly for all  $p, q$ . That is, the coefficient hidden by  $O$ -notation does not depend on  $p$  and  $q$ .

From a technical perspective our main contribution is combinatorial and can be expressed as a solution to a specific *set nucleation* task. In such a task we start with a collection of “atomic compounds” (a collection of singleton sets), and the goal is to assemble a specified collection of “target compounds” (a collection of sets that are unions of the singletons). The assembly is to be executed by a straight-line program, where each operation in the program selects two *disjoint* sets in the collection and inserts their union into the collection. (Once a set is in the collection, it may be selected arbitrarily many times.) The assembly should be done in as few operations as possible.

Our main contribution can be viewed as a straight-line program of length  $O((n^p + n^q) \log n)$  that assembles the collection  $\{\{X : X \cap Y = \emptyset\} : Y\}$  starting from the collection  $\{\{X\} : X\}$ , where  $X$  ranges over the  $p$ -subsets of  $[n]$  and  $Y$  ranges over the  $q$ -subsets of  $[n]$ . Valiant’s lemma [23] in these terms provides an optimal solution of length  $\Theta(n)$  for the specific case  $p = q = 1$ .

**1.5. Applications.** Many classical optimisation problems and counting problems can be algebrised over a commutative semigroup. A selection of applications will be reviewed in Sect. 3.

**1.6. Related work.** “Nucleation” is implicit in the design of many fast algebraic algorithms, perhaps two of the most central are the fast Fourier transform of Cooley and Tukey [12] (as is witnessed by the butterfly circuit representation) and Yates’s 1937 algorithm [27] for computing the product of a vector with the tensor product of  $n$  matrices of size  $2 \times 2$ . The latter can in fact be directly used to obtain a nucleation process for  $(p, q)$ -disjoint summation, even if an inefficient one. (For an exposition of Yates’s method we recommend Knuth [20, §4.6.4]; take  $m_i = 2$  and  $g_i(s_i, t_i) = [s_i = 0 \text{ or } t_i = 0]$  for  $i = 1, 2, \dots, n$  to extract the following nucleation process implicit in the algorithm.) For all  $Z \subseteq [n]$  and  $i \in \{0, 1, \dots, n\}$ , let

$$(2) \quad a_i(Z) = \{X \subseteq [n] : X \cap [n - i] = Z \cap [n - i], X \cap Z \setminus [n - i] = \emptyset\}.$$

Put otherwise,  $a_i(Z)$  consists of  $X$  that agree with  $Z$  in the first  $n - i$  elements of  $[n]$  and are disjoint from  $Z$  in the last  $i$  elements of  $[n]$ . In particular, our objective is to assemble the sets  $a_n(Y) = \{X : X \cap Y = \emptyset\}$  for each  $Y \subseteq [n]$  starting from the singletons  $a_0(X) = \{X\}$  for each  $X \subseteq [n]$ . The nucleation process given by Yates’ algorithm is, for all  $i = 1, 2, \dots, n$  and  $Z \subseteq [n]$ , to set

$$(3) \quad a_i(Z) = \begin{cases} a_{i-1}(Z \setminus \{n + 1 - i\}) & \text{if } n + 1 - i \in Z, \\ a_{i-1}(Z \cup \{n + 1 - i\}) \cup a_{i-1}(Z) & \text{if } n + 1 - i \notin Z. \end{cases}$$

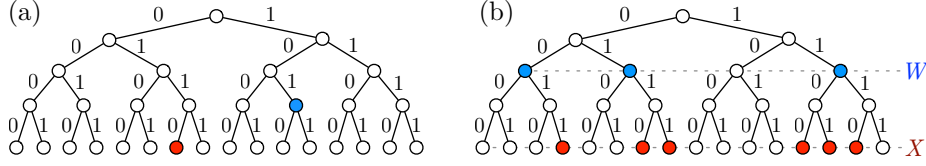


FIGURE 1. Representing  $\{0,1\}$ -strings of length at most  $b$  as nodes in a perfect binary tree of height  $b$ . Here  $b = 4$ . (a) Each string traces a unique path down from the root node, with the empty string  $\epsilon$  corresponding to the root node. The nodes at level  $0 \leq \ell \leq b$  correspond to the strings of length  $\ell$ . The red leaf node corresponds to 0110 and the blue node corresponds to 101. (b) A set of strings corresponds to a set of nodes in the tree. The set  $X$  is displayed in red, the set  $W$  in blue. The set  $W$  is the projection of the set  $X$  to level  $\ell = 2$ . Equivalently,  $X|_{\ell} = W$ .

This results in  $2^{n-1}n$  disjoint unions. If we restrict to the case  $|Y| \leq q$  and  $|X| \leq p$ , then it suffices to consider only  $Z$  with  $|Z| \leq p + q$ , which results in  $O((p + q) \sum_{j=0}^{p+q} \binom{n}{j})$  disjoint unions. Compared with our main result, this is not particularly efficient. In particular, our main result relies on “tree-projection” partitioning that enables a significant speedup over the “prefix-suffix” partitioning in (2) and (3).

We observe that “set nucleation” can also be viewed as a computational problem, where the output collection is given and the task is to decide whether there is a straight-line program of length at most  $\ell$  that assembles the output using (disjoint) unions starting from singleton sets. This problem is known to be NP-complete even in the case where output sets have size 3 [15, Problem PO9]; moreover, the problem remains NP-complete if the unions are not required to be disjoint.

## 2. A CIRCUIT FOR $(p, q)$ -DISJOINT SUMMATION

**2.1. Nucleation of  $p$ -subsets with a perfect binary tree.** Looking at Valiant’s circuit construction in the introduction, we observe that the left half of the circuit accumulates sums of variables (i.e., sums of 1-subsets of  $[n]$ ) along what is a perfect binary tree. Our first objective is to develop a sufficient generalisation of this strategy to cover the setting where each summand is indexed by a  $p$ -subset of  $[n]$  with  $p \geq 1$ .

Let us assume that  $n = 2^b$  for a nonnegative integer  $b$  so that we can identify the elements of  $[n]$  with binary strings of length  $b$ . We can view each binary string of length  $b$  as traversing a unique path starting from the root node of a perfect binary tree of height  $b$  and ending at a unique leaf node. Similarly, we may identify any node at level  $\ell$  of the tree by a binary string of length  $\ell$ , with  $0 \leq \ell \leq b$ . See Fig. 1(a) for an illustration. For  $p = 1$  this correspondence suffices.

For  $p > 1$ , we are not studying individual binary strings of length  $b$  (that is, individual elements of  $[n]$ ), but rather  $p$ -subsets of such strings. In particular, we can identify each  $p$ -subset of  $[n]$  with a  $p$ -subset of leaf nodes in the binary

tree. To nucleate such subsets it will be useful to be able to “project” sets upward in the tree. This motivates the following definitions.

Let us write  $\{0, 1\}^\ell$  for the set of all binary strings of length  $0 \leq \ell \leq b$ . For  $\ell = 0$ , we write  $\epsilon$  for the empty string. For a subset  $X \subseteq \{0, 1\}^b$ , we define the *projection of  $X$  to level  $\ell$*  as

$$(4) \quad X|_\ell = \left\{ x \in \{0, 1\}^\ell : \exists y \in \{0, 1\}^{b-\ell} \text{ such that } xy \in X \right\}.$$

That is,  $X|_\ell$  is the set of length- $\ell$  prefixes of strings in  $X$ . Equivalently, in the binary tree we obtain  $X|_\ell$  by lifting each element of  $X$  to its ancestor on level  $\ell$  in the tree. See Fig. 1(b) for an illustration. For the empty set we define  $\emptyset|_\ell = \emptyset$ .

Let us now study a set family  $\mathcal{F} \subseteq 2^{\{0,1\}^b}$ . The intuition here is that each member of  $\mathcal{F}$  is a summand, and  $\mathcal{F}$  represents the sum of its members. A circuit design must assemble (nucleate)  $\mathcal{F}$  by taking disjoint unions of carefully selected subfamilies. This motivates the following definitions.

For a level  $0 \leq \ell \leq b$  and a string  $W \subseteq \{0, 1\}^\ell$  let us define *the subfamily of  $\mathcal{F}$  that projects to  $W$*  by

$$(5) \quad \mathcal{F}_W = \{X \in \mathcal{F} : X|_\ell = W\}.$$

That is, the family  $\mathcal{F}_W$  consists of precisely those members  $X \in \mathcal{F}$  that project to  $W$ . Again Fig. 1(b) provides an illustration: we select precisely those  $X$  whose projection is  $W$ .

The following technical observations are now immediate. For each  $0 \leq \ell \leq b$ , if  $\emptyset \in \mathcal{F}$ , then we have

$$(6) \quad \mathcal{F}_\emptyset = \{\emptyset\}.$$

Similarly, for  $\ell = 0$  we have

$$(7) \quad \mathcal{F}_{\{\epsilon\}} = \mathcal{F} \setminus \{\emptyset\}.$$

For  $\ell = b$  we have for every  $W \in \mathcal{F}$  that

$$(8) \quad \mathcal{F}_W = \{W\}.$$

Now let us restrict our study to the situation where the family  $\mathcal{F} \subseteq 2^{\{0,1\}^b}$  contains only sets of size at most  $p$ . In particular, this is the case in our applications. For a set  $U$  and an integer  $p$ , let us write  $\binom{U}{p}$  for the family of all subsets of  $U$  of size  $p$ , and  $\binom{U}{\downarrow p}$  for the family of all subsets of  $U$  with size at most  $p$ . Accordingly, for integers  $0 \leq k \leq n$ , let us use the shorthand  $\binom{n}{\downarrow k} = \sum_{i=0}^k \binom{n}{i}$ .

The following lemma enables us to recursively nucleate any family  $\mathcal{F} \subseteq \binom{\{0,1\}^b}{\downarrow p}$ . In particular, we can nucleate the family  $\mathcal{F}_W$  with  $W$  in level  $\ell$  using the families  $\mathcal{F}_Z$  with  $Z$  in level  $\ell + 1$ . Applied recursively, we obtain  $\mathcal{F}$  by proceeding from the bottom up, that is,  $\ell = b, b-1, \dots, 1, 0$ . The intuition underlying the lemma is illustrated in Fig. 2.

**Lemma 1.** *For all  $0 \leq \ell \leq b-1$ ,  $\mathcal{F} \subseteq \binom{\{0,1\}^b}{\downarrow p}$ , and  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ , we have that the family  $\mathcal{F}_W$  is a disjoint union  $\mathcal{F}_W = \bigcup \left\{ \mathcal{F}_Z : Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p} \right\}_W$ .*

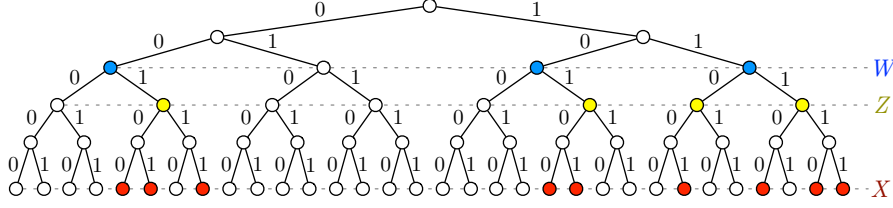


FIGURE 2. Illustrating the proof of Lemma 1. Here  $b = 5$ . The set  $X$  (indicated with red nodes) projects to level  $\ell = 2$  to the set  $W$  (indicated with blue nodes) and to level  $\ell + 1 = 3$  to the set  $Z$  (indicated with yellow nodes). Furthermore, the projection of  $Z$  to level  $\ell$  is  $W$ . Thus, each  $X \in \mathcal{F}$  is included to  $\mathcal{F}_W$  exactly from  $\mathcal{F}_Z$  in Lemma 1.

*Proof.* The projection of each  $X \in \mathcal{F}$  to level  $\ell + 1$  is unique, so the families  $\mathcal{F}_Z$  are pairwise disjoint for distinct  $Z$ . Now consider an arbitrary  $X \in \mathcal{F}$  and set  $X|_{\ell+1} = Z$ , that is,  $X \in \mathcal{F}_Z$ . From (4) we have  $X|_\ell = Z|_\ell$ , which implies that we have  $X \in \mathcal{F}_W$  if and only if  $X|_\ell = W$  if and only if  $Z|_\ell = W$  if and only if  $Z \in \left( \begin{smallmatrix} \{0,1\}^{\ell+1} \\ \downarrow p \end{smallmatrix} \right)_W$ .  $\square$

**2.2. A generalisation:  $(p, q)$ -intersection summation.** It will be convenient to study a minor generalisation of  $(p, q)$ -disjoint summation. Namely, instead of insisting on disjointness, we allow nonempty intersections to occur with “active” (or “avoided”)  $q$ -subsets  $A$ , but require that elements in the intersection of each  $p$ -subset and each  $A$  are “individualized.” That is, our input is not given by associating a value  $f(X) \in S$  to each set  $X \in \left( \begin{smallmatrix} [n] \\ \downarrow p \end{smallmatrix} \right)$ , but is instead given by associating a value  $g(I, X) \in S$  to each pair  $(I, X)$  with  $I \subseteq X \in \left( \begin{smallmatrix} [n] \\ \downarrow p \end{smallmatrix} \right)$ , where  $I$  indicates the elements of  $X$  that are “individualized.” In particular, we may insist (by appending to  $S$  a formal identity element if such an element does not already exist in  $S$ ) that  $g(I, X)$  vanishes unless  $I$  is empty. This reduces  $(p, q)$ -disjoint summation to the following problem:

**Problem 2. ( $(p, q)$ -intersection summation)** Given as input a function  $g$  that maps each pair  $(I, X)$  with  $I \subseteq X \in \left( \begin{smallmatrix} [n] \\ \downarrow p \end{smallmatrix} \right)$  and  $|I| \leq q$  to an element  $g(I, X) \in S$ , output the function  $h: \left( \begin{smallmatrix} [n] \\ \downarrow q \end{smallmatrix} \right) \rightarrow S$  defined for all  $A \in \left( \begin{smallmatrix} [n] \\ \downarrow q \end{smallmatrix} \right)$  by

$$(9) \quad h(A) = \bigoplus_{X \in \left( \begin{smallmatrix} [n] \\ \downarrow p \end{smallmatrix} \right)} g(A \cap X, X).$$

**2.3. The circuit construction.** We proceed to derive a recursion for the function  $h$  using Lemma 1 to carry out nucleation of  $p$ -subsets. The recursion proceeds from the bottom up, that is,  $\ell = b, b - 1, \dots, 1, 0$  in the binary tree representation. (Recall that we identify the elements of  $[n]$  with the elements of  $\{0, 1\}^b$ , where  $n$  is a power of 2 with  $n = 2^b$ .) The intermediate functions  $h_\ell$  computed by the recursion are “projections” of (9) using (5). In more precise terms, for  $\ell = b, b - 1, \dots, 1, 0$ , the function  $h_\ell: \left( \begin{smallmatrix} \{0,1\}^b \\ \downarrow q \end{smallmatrix} \right) \times \left( \begin{smallmatrix} \{0,1\}^\ell \\ \downarrow p \end{smallmatrix} \right) \rightarrow S$

is defined for all  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$  and  $A \in \binom{\{0,1\}^b}{\downarrow q}$  by

$$(10) \quad h_\ell(A, W) = \bigoplus_{X \in \binom{\{0,1\}^b}{\downarrow p}_W} g(A \cap X, X).$$

Let us now observe that we can indeed recover the function  $h$  from the case  $\ell = 0$ . Indeed, for the empty string  $\epsilon$ , the empty set  $\emptyset$  and every  $A \in \binom{\{0,1\}^b}{\downarrow q}$  we have by (6) and (7) that

$$(11) \quad h(A) = h_0(A, \{\epsilon\}) \oplus h_0(A, \emptyset).$$

It remains to derive the recursion that gives us  $h_0$ . Here we require one more technical observation, which enables us to narrow down the intermediate values  $h_\ell(A, W)$  that need to be computed to obtain  $h_0$ . In particular, we may discard the part of the active set  $A$  that extends outside the “span” of  $W$ . This observation is the crux in deriving a succinct circuit design.

For  $0 \leq \ell \leq b$  and  $w \in \{0,1\}^\ell$ , we define the *span* of  $w$  by

$$\langle w \rangle = \{x \in \{0,1\}^b : \exists z \in \{0,1\}^{b-\ell} \text{ such that } wz = x\}.$$

In the binary tree,  $\langle w \rangle$  consists of the leaf nodes in the subtree rooted at  $w$ . Let us extend this notation to subsets  $W \subseteq \{0,1\}^\ell$  by  $\langle W \rangle = \bigcup_{w \in W} \langle w \rangle$ . The following lemma shows that it is sufficient to evaluate  $h_\ell(A, W)$  only for  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$  and  $A \in \binom{\{0,1\}^b}{\downarrow q}$  such that  $A \subseteq \langle W \rangle$ .

**Lemma 3.** *For all  $0 \leq \ell \leq b$ ,  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ , and  $A \in \binom{\{0,1\}^b}{\downarrow q}$ , we have*

$$(12) \quad h_\ell(A, W) = h_\ell(A \cap \langle W \rangle, W).$$

*Proof.* If  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ ,  $X \in \binom{\{0,1\}^b}{\downarrow p}$ , and  $X|_\ell = W$ , then we have  $X \subseteq \langle W \rangle$ . Thus, directly from (10), we have

$$\begin{aligned} h_\ell(A, W) &= \bigoplus_{X \in \binom{\{0,1\}^b}{\downarrow p}_W} g(A \cap X, X) \\ &= \bigoplus_{X \in \binom{\{0,1\}^b}{\downarrow p}_W} g((A \cap \langle W \rangle) \cap X, X) \\ &= h_\ell(A \cap \langle W \rangle, W). \quad \square \end{aligned}$$

We are now ready to present the recursion for  $\ell = b, b-1, \dots, 1, 0$ . The base case  $\ell = b$  is obtained directly based on the values of  $g$ , because we have by (8) for all  $W \in \binom{\{0,1\}^b}{\downarrow p}$  and  $A \in \binom{\{0,1\}^b}{\downarrow q}$  with  $A \subseteq W$  that

$$(13) \quad h_b(A, W) = g(A, W).$$

The following lemma gives the recursive step from  $\ell + 1$  to  $\ell$  by combining Lemma 1 and Lemma 3.

**Lemma 4.** *For  $0 \leq \ell \leq b-1$ ,  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ , and  $A \in \binom{\{0,1\}^b}{\downarrow q}$  with  $A \subseteq \langle W \rangle$ , we have*

$$(14) \quad h_\ell(A, W) = \bigoplus_{Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W} h_{\ell+1}(A \cap \langle Z \rangle, Z).$$

*Proof.* We have

$$\begin{aligned}
h_\ell(A, W) &= \bigoplus_{X \in \binom{\{0,1\}^b}{\downarrow p}_W} g(A \cap X, X) & (10) \\
&= \bigoplus_{Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W} \bigoplus_{X \in \binom{\{0,1\}^b}{\downarrow p}_Z} g(A \cap X, X) & (\text{Lemma 1}) \\
&= \bigoplus_{Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W} h_{\ell+1}(A, Z) & (10) \\
&= \bigoplus_{Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W} h_{\ell+1}(A \cap \langle Z \rangle, Z). & (\text{Lemma 3}) \quad \square
\end{aligned}$$

The recursion given by (13), (14), and (12) now defines an arithmetic circuit that solves  $(p, q)$ -intersection summation.

**2.4. Size of the circuit.** By (13), the number of input gates in the circuit is equal to the number of pairs  $(I, X)$  with  $I \subseteq X \in \binom{\{0,1\}^b}{\downarrow p}$  and  $|X| \leq q$ , which is

$$(15) \quad \sum_{i=0}^p \sum_{j=0}^q \binom{2^b}{i} \binom{i}{j}.$$

To derive an expression for the number of  $\oplus$ -gates, we count for each  $0 \leq \ell \leq b-1$  the number of pairs  $(A, W)$  with  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ ,  $A \in \binom{\{0,1\}^b}{\downarrow q}$ , and  $A \subseteq \langle W \rangle$ , and for each such pair  $(A, W)$  we count the number of  $\oplus$ -gates in the subcircuit that computes the value  $h_\ell(A, W)$  from the values of  $h_{\ell+1}$  using (14).

First, we observe that for each  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$  we have  $|\langle W \rangle| = 2^{b-\ell} |W|$ . Thus, the number of pairs  $(A, W)$  with  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ ,  $A \in \binom{\{0,1\}^b}{\downarrow q}$ , and  $A \subseteq \langle W \rangle$  is

$$(16) \quad \sum_{i=0}^p \sum_{j=0}^q \binom{2^\ell}{i} \binom{i 2^{b-\ell}}{j}.$$

For each such pair  $(A, W)$ , the number of  $\oplus$ -gates for (14) is  $\left| \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W \right| - 1$ .

**Lemma 5.** For all  $0 \leq \ell \leq b-1$ ,  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$ , and  $|W| = i$ , we have

$$(17) \quad \left| \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W \right| = \sum_{k=0}^{p-i} \binom{i}{k} 2^{i-k}.$$

*Proof.* A set  $Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W$  can contain either one or both of the strings  $w0$  and  $w1$  for each  $w \in W$ . The set  $Z$  may contain both elements for at most  $p-i$  elements  $w \in W$  because otherwise  $|Z| > p$ . Finally, for each  $0 \leq k \leq p-i$ , there are  $\binom{i}{k} 2^{i-k}$  ways to select a set  $Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W$  such that  $Z$  contains  $w0$  and  $w1$  for exactly  $k$  elements  $w \in W$ .  $\square$



Finally, for each  $A \in \binom{\{0,1\}^b}{\downarrow q}$  we require an  $\oplus$ -gate that is also designated as an output gate to implement (11). The number of these gates is

$$(18) \quad \sum_{j=0}^q \binom{2^b}{j}.$$

The total number of  $\oplus$ -gates in the circuit is obtained by combining (15), (16), (17), and (18). The number of  $\oplus$ -gates is thus

$$\begin{aligned} & \sum_{i=0}^p \sum_{j=0}^q \binom{2^b}{i} \binom{i}{j} + \sum_{\ell=0}^{b-1} \sum_{i=0}^p \sum_{j=0}^q \binom{2^\ell}{i} \binom{i 2^{b-\ell}}{j} \left( \sum_{k=0}^{p-i} \binom{i}{k} 2^{i-k} - 1 \right) + \sum_{j=0}^q \binom{2^b}{j} \\ & \leq \sum_{\ell=0}^b \sum_{i=0}^p \sum_{j=0}^q \binom{2^\ell}{i} \binom{i 2^{b-\ell}}{j} 3^i \leq \sum_{\ell=0}^b \sum_{i=0}^p \sum_{j=0}^q \frac{(2^\ell)^i}{i!} \frac{i^j (2^{b-\ell})^j}{j!} 3^i \\ & \leq \sum_{\ell=0}^b \sum_{i=0}^p \sum_{j=0}^q \frac{(2^\ell)^{\max(p,q)}}{i!} \frac{i^j (2^{b-\ell})^{\max(p,q)}}{j!} 3^i \\ & = n^{\max(p,q)} (1 + \log_2 n) \sum_{i=0}^p \sum_{j=0}^q \frac{i^j 3^i}{i! j!}. \end{aligned}$$

The double sum is at most a constant because we have that

$$(19) \quad \sum_{i=0}^p \sum_{j=0}^q \frac{i^j 3^i}{i! j!} \leq \sum_{i=0}^{\infty} \frac{3^i}{i!} \sum_{j=0}^{\infty} \frac{i^j}{j!} = \sum_{i=0}^{\infty} \frac{3^i}{i!} e^i \leq \sum_{i=0}^{\infty} \frac{(3e^2)^i}{i^i},$$

where the last inequality follows from Stirling's formula. Furthermore,

$$(20) \quad \sum_{i=\lceil 6e^2 \rceil}^{\infty} \frac{(3e^2)^i}{i^i} \leq \sum_{i=\lceil 6e^2 \rceil}^{\infty} \frac{1}{2^i} \leq \sum_{i=0}^{\infty} \frac{1}{2^i} \leq 2.$$

Combining (19) and (20), we have

$$\sum_{i=0}^p \sum_{j=0}^q \frac{i^j 3^i}{i! j!} \leq \sum_{i=0}^{\lceil 6e^2 \rceil} \frac{(3e^2)^i}{i^i} + 2.$$

Thus, the circuit defined in §2 has size  $O((n^p + n^q) \log n)$ , where the constant hidden by the  $O$ -notation does not depend on  $p$  and  $q$ .

**2.5. Constructing the Circuit.** In this section we give an algorithm that outputs the circuit presented above, given  $b$ ,  $p$ , and  $q$  as input. This algorithm can also be used to compute (9) directly without constructing the circuit first.

**Algorithm 6.** Outputs a list of gates in the circuit, with labels on the input and output gates.

1. Initialise an associative data structure  $D$
2. For each  $X \in \binom{\{0,1\}^b}{\downarrow p}$  and  $I \in \binom{W}{\downarrow q}$ , create an input gate  $g$  labelled with  $(I, X)$  and set  $D(I, X) \leftarrow g$ .
3. Set  $\ell \leftarrow b - 1$ .
4. For each  $W \in \binom{\{0,1\}^\ell}{\downarrow p}$  and  $A \in \binom{W}{\downarrow q}$ ,
  - 4.1. select an arbitrary  $Z_0 \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W$  and set  $g \leftarrow D(A \cap \langle Z_0 \rangle, Z_0)$ ,

- 4.2. for each  $Z \in \binom{\{0,1\}^{\ell+1}}{\downarrow p}_W \setminus \{Z_0\}$  set  $g \leftarrow g \oplus D(A \cap \langle Z \rangle, Z)$ , and
- 4.3. set  $D(A, W) \leftarrow g$ .
5. If  $\ell \geq 1$ , set  $\ell \leftarrow \ell - 1$  and go to Step 4.
6. For each  $A \in \binom{\{0,1\}^b}{\downarrow q}$ , create an output gate  $D(A, \{\epsilon\}) \oplus D(A, \emptyset)$  labelled with  $A$ .

Letting  $k = \max(p, q)$ , the sets that appear in Algorithm 6 can be represented in  $O(k \log n)$  space, and each required operation on these sets can be done in  $O(k \log n)$  time. Thus, we observe that iterating over set families in Algorithm 6 takes  $O(k \log n)$  time per element, and the total number of iterations the algorithm makes is same as the number of gates in the circuit. Also, assuming that  $D$  is a self-balancing binary tree, each search and insert operation takes  $O((k \log n)^2)$  time, and a constant number of these operations is required for each gate. Thus, we have that the total running time of Algorithm 6 is  $O((p^2 + q^2)(n^p + n^q) \log^3 n)$ .

### 3. CONCLUDING REMARKS AND APPLICATIONS

We have generalised Valiant's [23] observation that negation is powerless for computing simultaneously the  $n$  different disjunctions of all but one of the given  $n$  variables: now we know that, in our terminology, subtraction is powerless for  $(p, q)$ -disjoint summation for any constant  $p$  and  $q$ . (Valiant proved this for  $p = q = 1$ .) Interestingly, requiring  $p$  and  $q$  be constants turns out to be essential, namely, when subtraction is available, an inclusion-exclusion technique is known [5] to yield a circuit of size  $O(p \binom{n}{\downarrow p} + q \binom{n}{\downarrow q})$ , which, in terms of  $p$  and  $q$ , is exponentially smaller than our bound  $O((n^p + n^q) \log n)$ . This gap highlights the difference of the algorithmic ideas behind the two results. Whether the gap can be improved to polynomial in  $p$  and  $q$  is an open question.

While we have dealt with the abstract notions of “monotone sums” or semigroup sums, in applications they most often materialise as maximisation or minimisation, as described in the next paragraphs. Also, in applications local terms are usually combined not only by one (monotone) operation but two different operations, such as “min” and “+”. To facilitate the treatment of such applications, we extend the semigroup to a semiring  $(S, \oplus, \odot)$  by introducing a product operation “ $\odot$ ”. Now the task is to evaluate

$$(21) \quad \bigoplus_{X, Y: X \cap Y = \emptyset} f(X) \odot g(Y),$$

where  $X$  and  $Y$  run through all  $p$ -subsets and  $q$ -subsets of  $[n]$ , respectively, and  $f$  and  $g$  are given mappings to  $S$ . We immediately observe that the expression (21) is equal to  $\bigoplus_Y e(Y) \odot g(Y)$ , where the sum is over all  $q$ -subsets of  $[n]$  and  $e$  is as in (1). Thus, by our main result, it can be evaluated using a circuit with  $O((n^p + n^q) \log n)$  gates.

**3.1. Application to  $k$ -paths.** We apply the semiring formulation to the problem of counting the maximum-weight  $k$ -edge paths from vertex  $s$  to vertex  $t$  in a given edge-weighted graph with real weights, where we assume that we are only allowed to add and compare real numbers and these operations take constant time (cf. [25]). By straightforward Bellman–Held–Karp type

dynamic programming [2, 3, 16] (or, even by brute force) we can solve the problem in  $\binom{n}{k}n^{O(1)}$  time. However, our main result gives an algorithm that runs in  $n^{k/2+O(1)}$  time by solving the problem in halves: Guess a middle vertex  $v$  and define  $f_1(X)$  as the number of maximum-weight  $k/2$ -edge paths from  $s$  to  $v$  in the graph induced by the vertex set  $X \cup \{v\}$ ; similarly define  $g_1(X)$  for the  $k/2$ -edge paths from  $v$  to  $t$ . Furthermore, define  $f_2(X)$  and  $g_2(X)$  as the respective maximum weights and put  $f(X) = (f_1(X), f_2(X))$  and  $g(X) = (g_1(X), g_2(X))$ . These values can be computed for all vertex subsets  $X$  of size  $k/2$  in  $\binom{n}{k/2}n^{O(1)}$  time. It remains to define the semiring operations in such a way that the expression (21) equals the desired number of  $k$ -edge paths; one can verify that the following definitions work correctly:  $(c, w) \odot (c', w') = (c \cdot c', w + w')$  and

$$(c, w) \oplus (c', w') = \begin{cases} (c, w) & \text{if } w > w', \\ (c', w') & \text{if } w < w', \\ (c + c', w) & \text{if } w = w'. \end{cases}$$

[[Please see Appendix B for details.]]

Thus, the techniques of the present paper enable solving the problem essentially as fast as the fastest known algorithms for the special case of counting *all* the  $k$ -paths, for which quite different techniques relying on subtraction yield  $\binom{n}{k/2}n^{O(1)}$  time bound [7]. On the other, for the more general problem of counting weighted subgraphs Vassilevska and Williams [24] give an algorithm whose running time, when applied to  $k$ -paths, is  $O(n^{\omega k/3 + n^{2k/3+c}})$ , where  $\omega < 2.3727$  is the exponent of matrix multiplication and  $c$  is a constant; this of course would remain worse than our bound even if  $\omega = 2$ .

**3.2. Application to matrix permanent.** Consider the problem of computing the permanent of a  $k \times n$  matrix  $(a_{ij})$  over a *noncommutative semiring*, with  $k \leq n$  and even for simplicity, given by  $\sum_{\sigma} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{k\sigma(k)}$ , where the sum is over all injective mappings  $\sigma$  from  $[k]$  to  $[n]$ . We observe that the expression (21) equals the permanent if we let  $p = q = k/2 = \ell$  and define  $f(X)$  as the sum of  $a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{\ell\sigma(\ell)}$  over all injective mappings  $\sigma$  from  $\{1, 2, \dots, \ell\}$  to  $X$  and, similarly,  $g(Y)$  as the sum of  $a_{\ell+1\sigma(\ell+1)} a_{\ell+2\sigma(\ell+2)} \cdots a_{k\sigma(k)}$  over all injective mappings  $\sigma$  from  $\{\ell+1, \ell+2, \dots, k\}$  to  $Y$ . Since the values  $f(X)$  and  $g(Y)$  for all relevant  $X$  and  $Y$  can be computed by dynamic programming in  $\binom{n}{k/2}n^{O(1)}$  time, our main result yields the time bound  $n^{k/2+O(1)}$  for computing the permanent.

Thus we improve significantly upon a Bellman–Held–Karp type dynamic programming algorithm that computes the permanent in  $\binom{n}{k}n^{O(1)}$  time, the best previous upper bound we are aware of for noncommutative semirings [8]. It should be noted, however, that essentially as fast algorithms are already known for *noncommutative rings* [8], and that faster,  $2^k n^{O(1)}$  time, algorithms are known for *commutative semirings* [8, 21].

**3.3. Application to feature selection.** The extensively studied feature selection problem in machine learning asks for a subset  $X$  of a given set of available features  $A$  so as to maximise some objective function  $f(X)$ . Often the size of  $X$  can be bounded from above by some constant  $k$ , and sometimes the selection task needs to be solved repeatedly with the set of available features  $A$  changing dynamically across, say, the set  $[n]$  of all features. Such constraints take place in a recent work [10] on Bayesian network structure learning by branch and bound: the algorithm proceeds by forcing some features,  $I$ , to be included in  $X$  and some other,  $E$ , to be excluded from  $X$ . Thus the key computational step becomes that of maximising  $f(X)$  subject to  $I \subseteq X \subseteq [n] \setminus E$  and  $|X| \leq k$ , which is repeated for varying  $I$  and  $E$ . We observe that instead of computing the maximum every time from scratch, it pays off precompute a solution to  $(p, q)$ -disjoint summation for all  $0 \leq p, q \leq k$ , since this takes about the same time as a single step for  $I = \emptyset$  and any fixed  $E$ . Indeed, in the scenario where the branch and bound search proceeds to exclude each and every subset of  $k$  features in turn, but no larger subsets, such precomputation decreases the running time bound quite dramatically, from  $O(n^{2k})$  to  $O(n^k)$ ; typically,  $n$  ranges from tens to some hundreds and  $k$  from 2 to 7. Admitted, in practice, one can expect the search procedure match the said scenario only partially, and so the savings will be more modest yet significant.

**Acknowledgement.** We thank Jukka Suomela for useful discussions.

## REFERENCES

- [1] Alon, N., Yuster, R., Zwick, U.: Finding and counting given length cycles. *Algorithmica* 17(3), 209–223 (1997)
- [2] Bellman, R.: Combinatorial processes and dynamic programming. In: *Combinatorial Analysis. Proceedings of Symposia in Applied Mathematics*, p. 217–249. No. 10, ACM (1960)
- [3] Bellman, R.: Dynamic programming treatment of the travelling salesman problem. *J. ACM* 9(1), 61–63 (1962)
- [4] Björklund, A.: Determinant sums for undirected Hamiltonicity. In: *51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010)*, pp. 173–182. IEEE Computer Society, Los Alamitos, CA (2010)
- [5] Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M.: Fourier meets möbius: fast subset convolution, manuscript submitted for publication
- [6] Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M.: Computing the Tutte polynomial in vertex-exponential time. In: *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pp. 677–686. IEEE Computer Society, Los Alamitos, CA (2008)
- [7] Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M.: Counting paths and packings in halves. In: *17th Annual European Symposium on Algorithms (ESA 2009)*. pp. 578–586. Springer, Berlin (2009)
- [8] Björklund, A., Husfeldt, T., Kaski, P., Koivisto, M.: Evaluation of permanents in rings and semirings. *Information Processing Letters* 110(20), 867–870 (2010)
- [9] Björklund, A., Husfeldt, T., Koivisto, M.: Set partitioning via inclusion-exclusion. *SIAM Journal on Computing* 39(2), 546–563 (2009)
- [10] de Campos, C.P., Ji, Q.: Efficient structure learning of bayesian networks using constraints. *Journal of Machine Learning Research* 12, 663–689 (2011)
- [11] Cohn, H., Kleinberg, R., Szegedy, B., Umans, C.: Group-theoretic algorithms for matrix multiplication. In: *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005)*. pp. 379–388. IEEE Computer Society, Los Alamitos, CA (2005)

- [12] Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* 19, 297–301 (1965)
- [13] Coppersmith, D., Winograd, S.: Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation* 9(3), 251–280 (1990)
- [14] Cygan, M., Nederlof, J., Pilipczuk, M., Pilipczuk, M., van Rooij, J.M.M., Wojtaszczyk, J.O.: Solving connectivity problems parameterized by treewidth in single exponential time (2011), manuscript
- [15] Garey, M.R., Johnson, D.S.: *Computers and Intractability, A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company (1979)
- [16] Held, M., Karp, R.M.: A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics* 10(1), 196–210 (1962)
- [17] Itai, A., Rodeh, M.: Finding a minimum circuit in a graph. *SIAM Journal on Computing* 7(4), 413–423 (1978)
- [18] Kaski, P., Koivisto, M., Korhonen, J.H.: Fast monotone summation over disjoint sets. In: 7th International Symposium on Parameterized and Exact Computation (IPEC 2012). Springer (2012), to appear.
- [19] Kerr, L.R.: The effect of algebraic structure on the computational complexity of matrix multiplications. Ph.D. thesis, Cornell University (1970)
- [20] Knuth, D.E.: *The Art of Computer Programming*, vol. 2, *Seminumerical Algorithms*. Addison–Wesley, Upper Saddle River, N.J., 3rd edn. (1998)
- [21] Koutis, I., Williams, R.: Limits and applications of group algebras for parameterized problems. In: *Automata, Languages and Programming, Lecture Notes in Computer Science*, vol. 5555, pp. 653–664. Springer, Berlin / Heidelberg (2009)
- [22] Lokshtanov, D., Nederlof, J.: Saving space by algebraization. In: 2010 ACM International Symposium on Theory of Computing (STOC 2010), pp. 321–330. ACM, New York (2010)
- [23] Valiant, L.G.: Negation is powerless for boolean slice functions. *SIAM Journal on Computing* 15, 531–535 (May 1986)
- [24] Vassilevska, V., Williams, R.: Finding, minimizing, and counting weighted subgraphs. In: 2009 ACM International Symposium on Theory of Computing (STOC 2009), pp. 455–464. ACM, New York (2009)
- [25] Vassilevska, V., Williams, R., Yuster, R.: Finding heaviest  $H$ -subgraphs in real weighted graphs, with applications. *ACM Transactions on Algorithms* 6(3), Art. 44, 23 (2010)
- [26] Williams, R.: Finding paths of length  $k$  in  $O^*(2^k)$  time. *Information Processing Letters* 109(6), 315–318 (2009)
- [27] Yates, F.: *The Design and Analysis of Factorial Experiments*. Imperial Bureau of Soil Science, Harpenden, England (1937)

## APPENDIX

### APPENDIX A. VALIANT'S CONSTRUCTION AND GENERALISATIONS

This section reviews Valiant's [23] circuit construction for  $(p, q)$ -disjoint summation in the case  $p = q = 1$ . We also present minor generalisations to the case when either  $p = 1$  or  $q = 1$ . For ease of exposition we follow the conventions and notation from Sect. 2. Accordingly, we assume that  $n = 2^b$  for a nonnegative integer  $b$  and identify the elements of  $[n]$  with binary strings in  $\{0, 1\}^b$ .

**A.1. Valiant's construction.** For  $p = q = 1$ , the disjoint summation problem reduces to the following form: given  $f: \{0, 1\}^b \rightarrow S$  as input, compute  $e: \{0, 1\}^b \rightarrow S$  defined for all  $y, x \in \{0, 1\}^b$  by

$$(22) \quad e(y) = \bigoplus_{x: x \neq y} f(x).$$

Valiant's construction computes these sums by first computing intermediate functions  $h_\ell^+: \{0, 1\}^\ell \rightarrow S$  defined for  $\ell = 1, 2, \dots, b$  and  $w \in \{0, 1\}^\ell$  by

$$(23) \quad h_\ell^+(w) = \bigoplus_{x \in \langle w \rangle} f(x),$$

and  $h_\ell^-: \{0, 1\}^\ell \rightarrow S$  defined for  $\ell = 1, 2, \dots, b$  and  $u \in \{0, 1\}^\ell$  by

$$(24) \quad h_\ell^-(u) = \bigoplus_{x \in \{0, 1\}^b \setminus \langle u \rangle} f(x).$$

The solution to (22) can then be recovered as  $e(y) = h_b^-(y)$  for all  $y \in \{0, 1\}^b$ .

The values (23) can be computed for  $\ell = b, b-1, \dots, 2, 1$  using the recurrence

$$(25) \quad \begin{aligned} h_b^+(x) &= f(x) \\ h_\ell^+(w) &= h_{\ell+1}^+(w0) \oplus h_{\ell+1}^+(w1). \end{aligned}$$

Assuming that functions  $h_\ell^+$  have been computed for all  $\ell$ , the values (24) can then be computed for  $\ell = 1, 2, \dots, b$  as

$$(26) \quad \begin{aligned} h_1^-(u) &= h_1^+(1-u) \\ h_\ell^-(ui) &= h_{\ell-1}^-(u) \oplus h_\ell^+(u(1-i)). \end{aligned}$$

The number of  $\oplus$ -gates to implement (25) and (26) as a circuit is exactly

$$\sum_{i=1}^{b-1} 2^i + \sum_{i=2}^b 2^i = 3n - 6 = O(n).$$

**A.2. Generalisation for  $p = 1$  and  $q > 1$ .** For  $p = 1$  and  $q > 1$ , the  $(p, q)$ -disjoint summation problem reduces to computing

$$(27) \quad e(Y) = \bigoplus_{x: x \notin Y} f(x),$$

where  $x \in \{0, 1\}^b$  and  $Y \in \binom{\{0, 1\}^b}{q}$ .

To evaluate (27), we proceed analogously to the  $p = q = 1$  case, first computing functions  $h_\ell^+: \{0, 1\}^\ell \rightarrow S$  as defined in (23). The second set of

intermediate functions now consists of functions  $h_\ell^-: \binom{\{0,1\}^\ell}{\downarrow q} \rightarrow S$ , defined for  $\ell = 0, 1, \dots, b$  and  $U \in \binom{\{0,1\}^\ell}{\downarrow q}$  by

$$(28) \quad h_\ell^-(U) = \bigoplus_{x \in \{0,1\}^b \setminus \langle U \rangle} f(x).$$

Then we have  $e(Y) = h_b^-(Y)$  for all  $Y \in \binom{\{0,1\}^b}{\downarrow q}$ .

The values  $h_\ell^+$  are computed as in Valiant's construction. For  $\ell = 0, 1, \dots, b$  and  $U \in \binom{\{0,1\}^\ell}{\downarrow q}$ , define

$$\hat{U} = \{x(1-i): xi \in U \text{ and } x(1-i) \notin U\}.$$

Now we can evaluate (28) using recurrence

$$(29) \quad \begin{aligned} h_0^-(\{\epsilon\}) &= 0 \\ h_0^-(\emptyset) &= h_1^+(0) \oplus h_1^+(1) \\ h_\ell^-(U) &= h_{\ell-1}^-(U|_{\ell-1}) \oplus \bigoplus_{x \in \hat{U}} h_\ell^+(x). \end{aligned}$$

The number of  $\oplus$ -gates to evaluate the values  $h_\ell^+$  is  $n - 2$ . To determine the number of  $\oplus$ -gates to evaluate the values  $h_\ell^-$ , we note that  $|\hat{U}| \leq |U|$ , and thus  $|\hat{U}| \leq q$  gates are used for any  $U \in \binom{\{0,1\}^\ell}{\downarrow q}$ . Thus, the total number of  $\oplus$ -gates is at most

$$1 + q \cdot \left| \bigcup_{\ell=1}^b \binom{\{0,1\}^\ell}{\downarrow q} \right| = 1 + q \sum_{\ell=1}^b \binom{2^\ell}{\downarrow q} = 1 + q \sum_{i=0}^q \sum_{\ell=1}^b \binom{2^\ell}{i}.$$

For positive integers  $i$  and  $\ell$  we have

$$\binom{2^{b-\ell}}{i} = \frac{2^b}{2^\ell i} \binom{2^{b-\ell}-1}{i-1} \leq \frac{2^b}{2^\ell i} \binom{2^b-1}{i-1} = \frac{1}{2^\ell} \binom{2^b}{i}.$$

Thus for positive integers  $i$  it holds that

$$\sum_{\ell=1}^b \binom{2^\ell}{i} = \sum_{\ell=0}^{b-1} \binom{2^{b-\ell}}{i} \leq \binom{2^b}{i} \sum_{\ell=0}^{b-1} \frac{1}{2^\ell} \leq 2 \binom{2^b}{i},$$

and hence

$$1 + q \sum_{i=0}^q \sum_{\ell=1}^b \binom{2^\ell}{i} \leq 1 + q \left( b + \sum_{i=1}^q 2 \binom{2^b}{i} \right) = O \left( q \binom{2^b}{\downarrow q} \right).$$

That is, the total number of  $\oplus$ -gates used is  $O(q \binom{n}{\downarrow q})$ .

**A.3. Generalisation for  $p > 1$  and  $q = 1$ .** For  $p > 1$  and  $q = 1$ , the  $(p, q)$ -disjoint summation problem reduces to computing

$$(30) \quad e(y) = \bigoplus_{X: y \notin X} f(X),$$

where  $X \in \binom{\{0,1\}^b}{p}$  and  $y \in \{0,1\}^b$ .

Now the first intermediate functions nucleate the inputs using tree-projection, that is, we define  $h_\ell^+ : \left(\begin{smallmatrix} \{0,1\}^\ell \\ \downarrow p \end{smallmatrix}\right) \rightarrow S$  for  $\ell = 0, 1, \dots, b$  and  $W \in \left(\begin{smallmatrix} \{0,1\}^\ell \\ \downarrow p \end{smallmatrix}\right)$  by

$$(31) \quad h_\ell^+(W) = \bigoplus_{X \in \left(\begin{smallmatrix} \{0,1\}^b \\ \downarrow p \end{smallmatrix}\right)_W} f(X).$$

We define the second intermediate functions  $h_\ell^- : \{0, 1\}^\ell \rightarrow S$  for all  $\ell = 0, 1, \dots, b$  and  $u \in \{0, 1\}^\ell$  by

$$(32) \quad h_\ell^-(u) = \bigoplus_{W: u \notin W} f(W),$$

where  $W$  ranges over  $\left(\begin{smallmatrix} \{0,1\}^\ell \\ \downarrow p \end{smallmatrix}\right)$ . Again  $e(y) = h_b^-(y)$  for all  $y \in \{0, 1\}^b$ .

By (8) and Lemma 1, we can compute (31) by the recurrence

$$\begin{aligned} h_b^+(X) &= f(X) \\ h_\ell^+(W) &= \bigoplus_{Z \in \left(\begin{smallmatrix} \{0,1\}^{\ell+1} \\ \downarrow p \end{smallmatrix}\right)_W} h_{\ell+1}^+(Z). \end{aligned}$$

Similarly, we can compute (32) by the recurrence

$$\begin{aligned} h_0^-(\{\epsilon\}) &= 0 \\ h_\ell^-(xi) &= h_{\ell-1}^-(x) \oplus \bigoplus_{W \in \left(\begin{smallmatrix} \{0,1\}^\ell \setminus \{x0, x1\} \\ \downarrow (p-1) \end{smallmatrix}\right)} h_\ell^+(\{x(1-i)\} \cup W). \end{aligned}$$

The number of  $\oplus$ -gates to evaluate  $h_\ell^+$  is  $\binom{2^b}{\downarrow p} - 1$ , and to evaluate  $h_\ell^-$  at most

$$\begin{aligned} 2 + \sum_{\ell=2}^b 2^\ell \binom{2^\ell - 2}{\downarrow (p-1)} &= 2 + \sum_{\ell=2}^b \sum_{i=0}^{p-1} 2^\ell \binom{2^\ell - 2}{i} \\ &\leq 2 + \sum_{\ell=2}^b \sum_{i=0}^{p-1} \frac{(i+1)2^\ell}{i+1} \binom{2^\ell - 1}{i} \\ &= 2 + \sum_{\ell=2}^b \sum_{i=0}^{p-1} (i+1) \binom{2^\ell}{i+1} \\ &\leq 2 + p \sum_{\ell=2}^b \binom{2^\ell}{\downarrow p} = O\left(p \binom{2^b}{\downarrow p}\right). \end{aligned}$$

Thus, the total number of  $\oplus$ -gates used is  $O(p \binom{n}{\downarrow p})$ .

**A.4. Generalisation for  $p > 1$  and  $q > 1$ .** It is an open problem whether a circuit of size  $O(p \binom{n}{\downarrow p} + q \binom{n}{\downarrow q})$  exists when  $p, q > 1$ . Our main result in Sect. 2 gives a circuit of size  $O((n^p + n^q) \log n)$ .

## APPENDIX B. COUNT-WEIGHT SEMIRINGS

This section reviews the count-weight semiring used in counting heaviest  $k$ -paths as described in §3.1. The following theorem is quite standard, but we give a detailed proof for the sake of completeness.



**Theorem 7.** *The Cartesian product  $\mathbb{N} \times (\mathbb{R} \cup \{-\infty\})$  is a commutative semiring when equipped with operations*

$$(33) \quad (c, w) \oplus (d, v) = \begin{cases} (c, w) & \text{if } w > v, \\ (d, v) & \text{if } w < v, \\ (c + d, w) & \text{if } w = v \end{cases}$$

and

$$(34) \quad (c, w) \odot (d, v) = (cd, w + v).$$

*Proof.* In the following, we will use the *Iverson bracket* notation; that is, if  $P$  is a predicate we have  $[P] = 1$  if  $P$  is true and  $[P] = 0$  if  $P$  is false. For the maximum of two elements  $x, y \in \mathbb{R}$ , we write  $x \vee y = \max(x, y)$ . We note that  $\vee$  is an associative, commutative binary operation on  $\mathbb{R} \cup \{-\infty\}$ .

First, let us note that it follows directly from (33) that

$$(c, w) \oplus (d, v) = (c[w = w \vee v] + d[v = w \vee v], w \vee v).$$

We now prove the claim using this observation and the known properties of  $+$  and  $\vee$ .

*Associativity of  $\oplus$ .* The associativity of  $\oplus$  follows from the associativity of  $+$  and  $\vee$ , as we have

$$\begin{aligned} & (c_1, w_1) \oplus ((c_2, w_2) \oplus (c_3, w_3)) \\ &= (c_1, w_1) \oplus (c_2[w_2 = w_2 \vee w_3] + c_3[w_3 = w_2 \vee w_3], w_2 \vee w_3) \\ &= \left( \sum_{i=1}^3 c_i [w_i = w_1 \vee (w_2 \vee w_3)], w_1 \vee (w_2 \vee w_3) \right) \\ &= \left( \sum_{i=1}^3 c_i [w_i = (w_1 \vee w_2) \vee w_3], (w_1 \vee w_2) \vee w_3 \right) \\ &= (c_1[w_1 = w_1 \vee w_2] + c_2[w_2 = w_1 \vee w_2], w_1 \vee w_2) \oplus (c_3, w_3) \\ &= ((c_1, w_1) \oplus (c_2, w_2)) \oplus (c_3, w_3). \end{aligned}$$

*Commutativity of  $\oplus$ .* The commutativity of  $\oplus$  follows from the commutativity of  $+$  and  $\vee$ , as we have

$$\begin{aligned} (c, w) \oplus (d, v) &= (c[w = w \vee v] + d[v = w \vee v], w \vee v) \\ &= (d[v = w \vee v] + c[w = w \vee v], w \vee v) \\ &= (d, v) \oplus (c, w). \end{aligned}$$

*Existence of additive identity.* We have that  $(0, -\infty)$  is the identity element for  $\oplus$ , as we have

$$\begin{aligned} (c, w) \oplus (0, -\infty) &= (c[w = w \vee -\infty] + 0[-\infty = w \vee -\infty], w \vee -\infty) \\ &= (c, w). \end{aligned}$$

*Associativity of  $\odot$ .* We have

$$\begin{aligned}
(c, w) \oplus ((d, v) \oplus (e, u)) \\
&= (c, w) \oplus (de, v + u) \\
&= (c(de), w + (v + u)) \\
&= ((cd)e, (w + v) + u) \\
&= (cd, w + v) \oplus (e, u) \\
&= ((c, w) \oplus (d, v)) \oplus (e, u).
\end{aligned}$$

*Commutativity of  $\odot$ .* We have

$$(c, w) \odot (d, v) = (cd, w + v) = (dc, v + w) = (d, v) \odot (c, w).$$

*Existence of multiplicative identity.* The multiplicative identity element is  $(1, 0)$ , since

$$(c, w) \odot (1, 0) = (c1, w + 0) = (c, w).$$

*Distributivity.* As  $\odot$  is commutativity, it suffices to prove that multiplication from left distributes over addition. We have

$$\begin{aligned}
(c, w) \odot ((d, v) \oplus (e, u)) \\
&= (c, w) \odot (d[v = v \vee u] + e[u = v \vee u], v \vee u) \\
&= (c(d[v = v \vee u] + e[u = v \vee u]), w + (v \vee u)) \\
&= (cd[v = v \vee u] + ce[u = v \vee u], (v + w) \vee (u + w)) \\
&= (cd[w + v = (w + v) \vee (w + u)] + ce[w + u = (w + v) \vee (w + u)], \\
&\quad (w + v) \vee (w + u)) \\
&= (cd, w + v) \oplus (ce, w + u) \\
&= ((c, w) \odot (d, v)) \oplus ((c, w) \odot (e, u)).
\end{aligned}$$

*Annihilation in multiplication.* Finally, we have that the additive identity element annihilates in multiplication, that is,

$$(c, w) \odot (0, -\infty) = (c0, w + -\infty) = (0, -\infty).$$

□